Semantic Segmentation of Building Models with Deep Learning in CityGML

Hanis Rashidan [1], Ivin Amri Musliman [1], Alias Abdul Rahman [1], Volker Coors [2], Gurcan Buyuksalih [3]

[1] 3D GIS Research Lab, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor Bahru, Malaysia - mhanis27@gmail.com, ivinamri@utm.my, alias@utm.my
[2] Centre for Geodesy and Geoinformatics, Stuttgart University of Applied Sciences (HFT Stuttgart), Stuttgart, Germany - volker.coors@hft-stuttgart.de
[3] Institute of Marine Sciences and Management, Istanbul University, Istanbul, Turkiye - gurcanb@istanbul.edu.tr

Keywords: 3D Buildings Models, Semantic Segmentation, CityGML.

Abstract

Semantic segmentation of 3D urban environments plays an important role in urban planning, management, and analysis. This paper presents an exploration of leveraging BuildingGNN, a deep learning framework for semantic segmentation of 3D building models, and the subsequent conversion of semantic labels into CityGML, the standardized format for 3D city models. The study begins with a methodology outlining the acquisition of a labelled dataset from BuildingNet and the necessary preprocessing steps for compatibility with BuildingGNN's architecture. The training process involves deep learning techniques tailored for 3D building structures, yielding insights into model performance metrics such as Intersection over Union (IoU) for several architectural components. Evaluation of the trained model highlights its accuracy and reliability, albeit with challenges observed, particularly in segmenting certain classes like doors. Moreover, the conversion of semantic labels into CityGML format is discussed, emphasizing the importance of data quality and meticulous annotation practices. The experiment as described in the methodology shows that outputs from the BuildingGNN for semantic segmentation can be utilized for the generation of CityGML building elements with some percentage of success. This particular work reveals several challenges such as the identification of individual architectural elements based on geometry groups. We believe that the improvement of the segmentation process could be further investigated in our near future work.

1. Introduction

As urban environments grow in complexity, the demand for detailed and accurate 3D city models represented in CityGML format becomes increasingly crucial for applications in urban planning, GIS, and smart city initiatives. CityGML, an open data model used for the representation and exchange of 3D city models in a standardized way (Kutzner et al., 2020), has prompted various methodologies in the generation of CityGML data. For instance, Voitenko (2023) explored the utilization of building footprints for CityGML generation, while an alternative approach involves the integration of BIM data into CityGML. This method requires geometry conversion and semantic mapping, as discussed by Biljecki et al. (2021) and Tan et al. (2023). Investigating the modelling of CityGML from LiDAR point cloud data, Jayaraj et al. (2018) contribute to the diverse landscape of data generation methods. Additionally, Biljecki et al. (2015) explore the conversion of OBJ models into CityGML. Despite the variety of approaches, challenges persist in the current ecosystem for creating and editing CityGML data. According to Jang et al. (2021), there are limitations in constructing CityGML data on a large scale with existing tools. Furthermore, Tan et al. (2023) highlight the deficiency in semantic mapping capabilities of current tools, impeding the complete conversion of CityGML data in practical applications. In response, this study explores an alternative approach that leverages deep learning for semantic segmentation in generating CityGML models from 3D model data.

Deep learning, a subset of machine learning, has demonstrated remarkable capabilities in analysing complex data structures, particularly in the realm of computer vision. Applied to the field of 3D data, deep learning methods have proven effective in tasks such as 3D semantic segmentation, where the goal is to assign meaningful labels to individual components of a three-dimensional scene. The motivation for semantic modelling in 3D data lies in promoting a structured approach to data representation, ensuring consistency, and fostering interoperability across diverse datasets (Uceda-Sosa et al., 2011). This involves creating models that not only capture the physical characteristics of buildings but also incorporate meaningful information about their components, functionalities, and relationships.

Research into the segmentation of building components within 3D models has explored various methodologies and technological approaches. Notable studies by Alexander and Ben (2015) and Hu et al. (2021) delve into the semantic segmentation of urban scenes, particularly focusing on point cloud data and point cloud classification. Meanwhile, Kundu et al. (2020) proposed a multiview fusion method, demonstrating its ability to achieve significantly better 3D semantic segmentation for indoor models. Building upon these advancements, Selvaraju et al. (2021) introduced a mesh-based graph neural network approach, leveraging modern deep backbones to automate the labelling of 3D building meshes.

In this context, it's crucial to emphasize the gap in the existing literature. The literature highlights the need for improved methodologies for generating CityGML data, particularly with regards to semantic mapping and scalability. However, while deep learning has shown potential in other domains, its application to semantic segmentation in CityGML generation remains relatively unexplored. Thus, this study aims to explore the gap by investigating the efficacy of deep learning techniques for semantic segmentation in the context of CityGML generation. Specifically, we seek to evaluate the performance of deep learning models in accurately segmenting building components from 3D model data and assess their potential for enhancing the generation of semantic-rich CityGML models.

This contribution has been peer-reviewed.

Figure 1 illustrates the overall process flow from 3D building models to a CityGML model, with the core segmentation process at the centre.

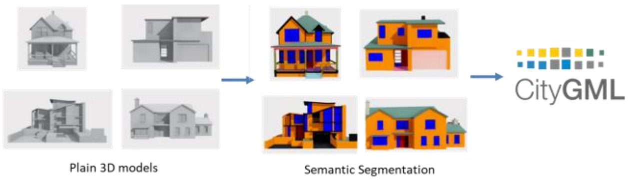

Figure 1. The overall flow of the building segmentation within CityGML.

The remainder of the paper describes the methodology in Section 2, discusses the findings in Section 3, and presents the conclusion in Section 4.

2. Methodology

This section describes the methodology for semantic segmentation of 3D building data by utilizing graph neural networks (GNN) and converting them into 3D GIS data standard such as CityGML. The focus is to segmentize the building exteriors with deep learning, providing a detailed representation of the model's structure. Figure 2 shows the processes involved in semantic segmentation - dataset, data preprocessing, AI development, inference, conversion to CityGML, and validation.

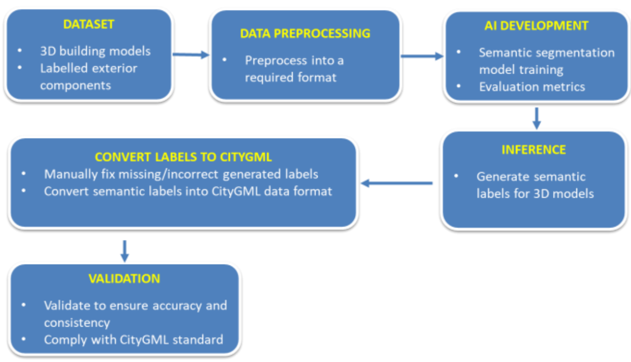

Figure 2. The process of semantic segmentation of 3D building data into CityGML.

2.1 Semantic Segmentation Through Deep Learning

Semantic segmentation in 3D models involves classifying different parts or components within a three-dimensional space. Unlike traditional pixel-based segmentation, this approach works directly on mesh data or point clouds, contributing to a more detailed understanding of the structural composition of urban environments. In adapting semantic segmentation to 3D models, various algorithms have been proposed to address the complexities of spatial data. These algorithms often incorporate graph neural networks (GNNs) and other deep learning techniques to effectively segment mesh data or point clouds. Examples include PointNet++ (Qi et al., 2017), MeshCNN (Hanocka et al., 2019), and MinkNet (Choy et al., 2019), each designed to capture and interpret the spatial relationships within 3D structures. The application of semantic segmentation to 3D models has potential for GIS, especially in the context of 3D building models conforming to the CityGML standard. This technique would enable the annotation and classification of distinct components within urban landscapes, facilitating a detailed representation of buildings and their features.

Current studies have investigated the segmentation of point clouds for urban feature extraction, but there is a notable gap in the application of advanced segmentation techniques to enrich the semantic content of 3D building models within CityGML standards. The advancement of semantic segmentation through deep learning offers prospects for enhancing the semantic content of 3D building models. By automating the segmentation of 3D building structures, this study aims to contribute to the development of enriched CityGML data. In this study, the labelling of semantic components involves the implementation of a specialized Graph Neural Network (GNN) known as BuildingGNN (Selvaraju et al., 2021). BuildingGNN serves as a tool for semantic segmentation, demonstrating its ability to label building models through an intricate analysis of spatial and structural relations among geometric primitives. The efficacy of BuildingGNN in the semantic labelling of 3D building data has been documented in research literature. Its notable performance has sparked interest for further research endeavours within the domain. The BuildingGNN approach operates by segmenting distinct "subgroups" within a building, treating each subgroup as a foundational component (i.e. windows, doors, roofs, and walls). The approach generates detailed representations for individual building components and establishes connections between the components using an edge-based approach (Selvaraju et al., 2021).

2.2 Semantic Segmentation Model Development

An essential aspect of developing the semantic segmentation model involves acquiring a quality labelled dataset. The dataset chosen for the model training comes from the BuildingNet dataset, and accessible at https://buildingnet.org. This dataset functions as a comprehensive repository of 3D building models, each uniformly labelled with exterior annotations for various architectural components. BuildingNet exhibits diversity, encompassing a range of architectural styles, sizes, and complexities. As detailed by Selvaraju et al. (2021), this dataset, methodically curated from Trimble's 3D Warehouse, results from a combination of crowdsourcing and expert guidance, yielding 513,000 annotated mesh primitives grouped into 292,000 semantic part components across 2,000 diverse building models.

2.2.1 Data Preprocessing

Given our utilization of the BuildingGNN framework, which employs the Minkowski Engine for efficient processing of sparse tensors in 3D space, preprocessing the dataset is essential. This ensures compatibility with the Minkowski network architecture. The Minkowski Engine excels at handling irregular, sparse data structures commonly encountered in 3D semantic segmentation tasks. In preparation for the Minkowski network, we performed preprocessing on the BuildingNet dataset. This involved organizing the data into training, testing, and validation subsets, as well as formatting appropriately for the consumption of the Minkowski network. This step ensures that the dataset aligns with the BuildingGNN’s input pipeline, which operates optimally with specific data structures and formats. This preprocessing not only streamlines the data ingestion process but also lays the groundwork for seamless integration with the BuildingGNN architecture, maximizing its effectiveness in semantic segmentation tasks.

2.2.2 Training Model Configuration

The training of the BuildingGNN model is organized by a shell script that encapsulates essential configurations crucial for model convergence and performance optimization. The script specifies key parameters, including the number of epochs set to 200 for comprehensive training. The model architecture employed is a variant of the Residual U-Net with 34 convolutional layers Res16UNet34C. Batch size is set to 32 to balance computational efficiency and memory utilization. Particularly, the script leverages CUDA-compatible GPUs to accelerate model training, ensuring expedited convergence. Additionally, the training process does not involve pretraining, suggesting that the model is trained from scratch using the specified dataset and configurations.

2.2.3 Evaluation Metrics

BuildingGNN introduces innovative evaluation methodologies tailored specifically for the semantic segmentation of 3D building structures, aiming to provide a comprehensive assessment of model performance. In the mesh track, traditional evaluation metrics are adapted to accommodate triangular meshes. BuildingGNN proposes modified Intersection over Union (IoU) variations that consider the contribution of each triangle, weighted by its face area, to offer a nuanced evaluation of segmentation accuracy.

In the evaluation process, BuildingGNN employs three primary metrics: part IoU, shape IoU, and per-triangle classification accuracy. Part IoU evaluation involves utilizing all annotated triangles across the test dataset, integrating both predicted and ground-truth labels. By incorporating triangle face areas into the IoU calculation, BuildingGNN aims to capture the spatial significance of individual triangles, providing insights into part-level segmentation accuracy. The shape IoU metric extends this evaluation to entire shapes, considering the diverse label distributions within annotated shapes.

Additionally, BuildingGNN reports per-triangle classification accuracy, acknowledging the impact of triangle size on classification performance. This evaluation framework enhances the interpretability of model results and facilitates a deeper understanding of semantic segmentation performance in complex 3D environments.

2.3 Converting Semantic Labels to CityGML

CityGML, designated as the standard data format for 3D city models, provides a robust framework for organizing and exchanging urban information. In our research, we leverage the capabilities of BuildingGNN to convert semantic labels into CityGML representations. This process plays a crucial role in integrating semantic information into standardized 3D city models, enhancing their utility for various urban applications.

While BuildingGNN encompasses over 20 labelled classes, our current exploration narrows down to a subset of CityObjects and thematic surfaces within the CityGML schema. Specifically, we concentrate on five primary classes: roof, wall, window, door, and ground surfaces. This selective approach allows us to delve into initial experimentation and explore the integration of semantic information within CityGML representations.

As part of the processing pipeline, BuildingGNN generates predictions in the form of labelled outputs stored in NPZ format. These outputs contain essential information, including predicted labels and per-triangle identifiers. This detailed data enables thorough analysis and evaluation of semantic segmentation results. To seamlessly integrate these outputs into CityGML representations using FME, we employ a conversion step that transforms NPZ files into JSON format. This conversion ensures efficient data processing and compatibility with downstream workflow.

Figure 3 illustrates the implementation of the conceptual mapping facilitated by FME, comprising several detailed steps for data processing and conversion into CityGML. The process begins with loading the labeled outputs in JSON format generated by BuildingGNN. These outputs contain valuable information about building features such as doors, windows, and walls. Once the data is loaded, a geometry filter is applied within the FME workspace. This filter categorizes features based on their geometry type, ensuring that each feature is accurately identified according to its geometric properties. Following this, an attribute filter is applied based on the class label IDs to further organize the data.

Subsequently, CityGML geometry and attribute property setters are utilized to format the data according to CityGML standards. This step involves mapping the semantic labels to the corresponding CityGML classes and attributes, ensuring coherence and alignment with CityGML's structured data model. This mapping is crucial for maintaining the integrity and usability of the data in CityGML format. Finally, a CityGML data writer completes the process by generating a CityGML file. For this work, CityGML version 2.0 is chosen, with LOD3 multi-surface representation to support the creation of detailed building features, including openings such as windows and doors. This version and level of detail are selected to ensure the accurate and detailed representation of the building models in the CityGML format.

By harnessing BuildingGNN’s capabilities, we establish a foundation for future endeavours. Our iterative approach aims to enrich CityGML representations with a comprehensive range of semantic attributes. This not only facilitates efficient data conversion but also deepens our understanding of the potential applications and implications of semantic urban modelling within the CityGML framework.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W11-2024-97-2024

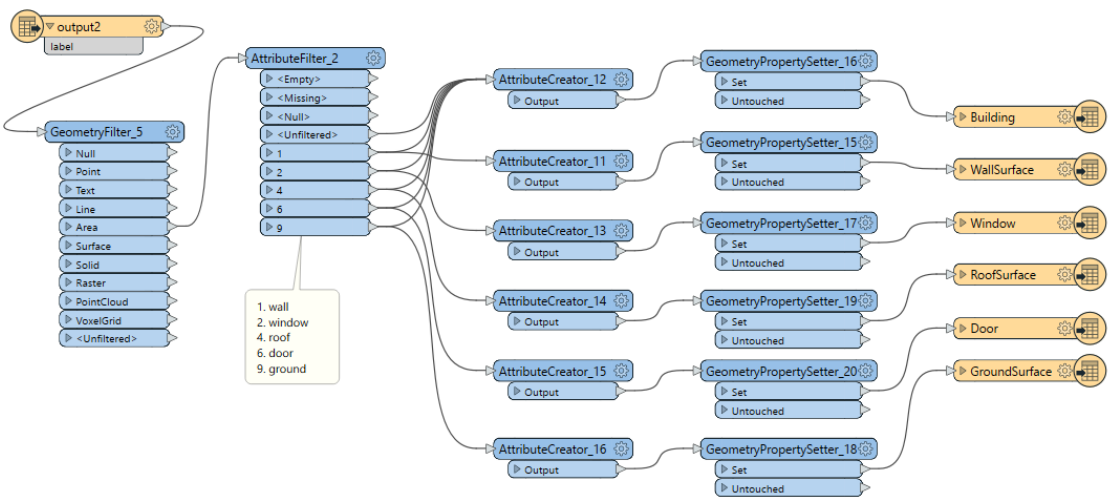

Figure 3. Implementation of the conceptual mapping from the generated labels to CityGML.

3. Discussions

In this study, the utilization of BuildingGNN for semantic segmentation of 3D building models, coupled with the conversion of semantic labels into the CityGML format, has provided insights into the accuracy and reliability of the process. The application of deep learning techniques, specifically tailored for 3D building structures, has demonstrated the ability of BuildingGNN to label architectural components.

3.1 Evaluation of Semantic Segmentation

Initial evaluation of the model training reveals the use of the IoU metric to assess segmentation performance. Table 1 presents the segmentation performance of the trained model based on triangle meshes. Notably, the highest accuracy was achieved for the "Ground" class, while the accuracy varied for other classes such as Roof, Wall, Window, and Door.

Classes	**IoU**
Roof	88.92
Wall	76.13
Window	72.05
Door	40.28
Ground	95.37

Table 1. Evaluation metrics of each label class.

Meanwhile, Figure 4 visually represents the final output of the semantic conversion process, depicting CityGML building models. Following the initial automatic generation of labels by the GNN models, manual corrections were implemented to address instances of incomplete or incorrect labelling. This iterative refinement process ensures the accuracy and completeness of the final semantic labels. In the rendered CityGML models, distinct colours are employed to represent various architectural components. The green delineates roof surfaces, while shades of blue depict window elements. The brown tones signify wall structures within the urban landscape.

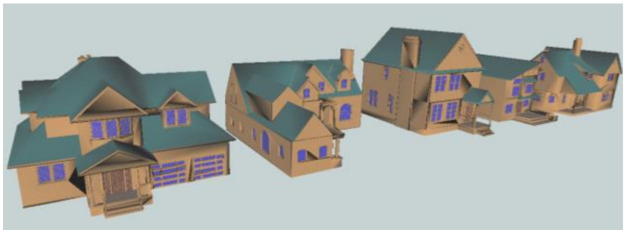

Figure 4. 3D models of CityGML with embedded semantic information.

The lower accuracy observed for the "Door" class in semantic segmentation can be attributed to several factors inherent to the complexities of urban environments and the nuances of deep learning models. Firstly, class imbalance within the training dataset plays a role, as datasets often exhibit variations in the distribution of labeled classes. In the case of doors, there are fewer instances of annotated doors compared to other classes, so the model may not receive sufficient exposure to effectively learn the distinguishing features of doors. Figure 5 shows a close-up view of one of the 3D building models, highlighting the similar shapes of windows and doors.

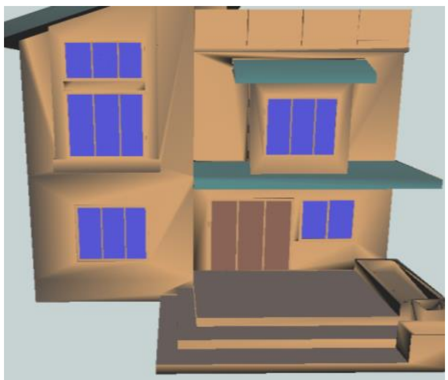

Figure 5. Close-up view of the 3D building models.

Moreover, the inherent complexity and variability of doors in urban settings pose significant challenges for correct segmentation. Doors come in various shapes, sizes, orientations, and contexts, rendering them highly diverse architectural elements. This variability introduces considerable ambiguity and intricacy into the segmentation task, making it difficult for the model to generalize effectively across different door types and conditions. We believe the preceding discussion warrants us for the challenges in developing a reliable and complete process for the segmentation.

3.2 The Training Data Quality Challenges

The quality of training data significantly impacts model performance. In our observations, we have encountered instances of inaccurate or inconsistent annotations. These issues include mislabeling and missing components, which can propagate errors throughout the training process. Ultimately, such issues undermine the model's ability to accurately discern and classify doors.

To address these challenges effectively, consider implementing robust annotation guidelines. Provide clear definitions and examples for each class to ensure a shared understanding among annotators. A standardized annotation protocol, complete with specific guidelines for labeling criteria and tools, can help maintain data quality. Keep in mind that while these measures may require additional time, cost, and resources, they are critical for ensuring reliable and effective models in real-world applications.

Furthermore, the architectural complexities of doors—characterized by detailed designs, handles, and textures—pose additional challenges for segmentation models. Exploring advanced model architectures and strategies to augment model capacity can potentially enhance performance in handling these complexities.

3.3 Downsampling of 3D Models

Downsampling of 3D data is a common preprocessing step in semantic segmentation tasks. This process involves reducing the spatial resolution of the input data. The primary motivation behind downsampling is to manage the computational complexity associated with processing high-resolution 3D data. By aggregating information from neighboring geometries or points, downsampling achieves a more balanced representation of the input data, thereby enhancing the model's ability to generalize.

However, downsampling introduces trade-offs. The reduction in spatial resolution leads to a loss of detail in the data, which can impact the model's performance - especially in tasks that require precise localization or identification of fine-grained features. Additionally, downsampling may introduce spatial aliasing artifacts. These artifacts occur when high-frequency components of the data are incorrectly represented at lower resolutions, potentially leading to inaccuracies in segmentation results.

When implementing downsampling as part of the preprocessing pipeline, careful consideration is essential. A balance between computational efficiency and preservation of relevant information is crucial. Optimizing downsampling parameters, such as the size or sampling rate, helps minimize information loss while still acquiring computational benefits.

3.4 CityGML Conversion Challenges

During the process of converting labeled data to CityGML format, a notable challenge emerges - specifically, the identification of architectural elements such as windows, doors, and roofs. Currently, this identification relies on the presence of specific geometry groups within the input mesh model. These geometry groups act as markers or indicators for different architectural components. However, when this information is missing or incomplete in the input data, it poses significant hurdles during the conversion process. As a result, the resulting CityGML models may lack the desired level of detail and specificity, impacting their usability and interoperability in urban modeling applications.

To address this challenge effectively, consider employing strategies for robust geometry detection and classification within the input mesh models. Techniques such as geometric pattern recognition, feature extraction, or rule-based inference can be leveraged to infer the presence and characteristics of architectural elements based on geometric properties and spatial relationships within the mesh data. By overcoming these challenges, the conversion of labeled data to CityGML can yield more accurate and semantically rich representations of urban environments. This, in turn, facilitates improved analysis, visualization, and decision-making in urban planning and management.

4. Conclusion and Near Future Work

In summary, this study has demonstrated the potential of utilizing deep learning for semantic segmentation and its integration with CityGML. However, critical challenges and opportunities lie ahead for the field. Our current findings represent only the initial steps in a complex journey toward more effective and reliable labeling processes for semantic urban data. While deep learning holds promise, it is not yet fully ready for real-world applications.

One major concern is the high tendency for mislabeling, which necessitates additional effort to correct. Manual intervention may be required to ensure effectiveness, adding extra time and resources to the process. To enhance our work, we can explore incorporating georeferenced objects, making it more useful for Geographic Information Systems (GIS) purposes.

Looking ahead, it is crucial to critically assess the limitations and shortcomings of existing segmentation models. Complex architectural designs, varying data quality, and the inherent ambiguity of urban scenes pose significant challenges that demand deeper exploration. Future research efforts may focus on developing more robust and adaptable segmentation algorithms.

Furthermore, automating the conversion pipeline process presents its own set of challenges. While advances in machine learning and geospatial technologies offer promising opportunities, seamless integration with CityGML remains complex. Addressing issues such as data interoperability, standardization of semantic labels, and scalability of conversion pipelines requires diligent attention.

References

Alexander Toshev and Ben Taskar, 2015. 3D all the way: Semantic segmentation of urban scenes from start to end in 3D. *In Proc. CVPR.*

Biljecki, F., Lim, J., Crawford, J., Moraru, D., Tauscher, H., Konde, A., Adouane, K., Lawrence, S., Janssen, P. and Stouffs, R., 2021. Extending CityGML for IFC-sourced 3D city models. *Automation in Construction*, 121, 103440.

Biljecki, F., & Ohori, K. A., 2015. Automatic Semantic-preserving Conversion Between OBJ and CityGML. *In UDMV* (pp. 25-30).

Choy, C., Gwak, J., & Savarese, S., 2019. 4D spatio-temporal convnets: Minkowski convolutional neural networks. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3075-3084).

Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., & Cohen-Or, D., 2019. Meshcnn: a network with an edge. *ACM Transactions on Graphics (ToG)*, 38(4), 1-12.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N., & Markham, A., 2021. Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4977-4987).

Jang, H., Kim, H., & Kang, H., 2021. Building large-scale CityGML feature for digital 3D infrastructure. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 39(3), 187-201.

Jayaraj, P., & Ramiya, A. M., 2018. 3D CityGML building modelling from lidar point cloud data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 175-180.

Kundu, A., Yin, X., Fathi, A., Ross, D., Brewington, B., Funkhouser, T., & Pantofaru, C., 2020. Virtual multi-view fusion for 3D semantic segmentation. *In Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16 (pp. 518-535). Springer International Publishing.

Kutzner, T., Chaturvedi, K., & Kolbe, T. H., 2020. CityGML 3.0: New functions open up new applications. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1), 43-61.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Selvaraju, P., Nabail, M., Loizou, M., Maslioukova, M., Averkiou, M., Andreou, A., Chaudhuri, S. and Kalogerakis, E., 2021. BuildingNet: Learning to label 3D buildings. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10397-10407).

Tan, Y., Liang, Y., & Zhu, J., 2023. CityGML in the Integration of BIM and the GIS: Challenges and Opportunities. *Buildings*, 13(7), 1758.

Uceda-Sosa, R., Srivastava, B., & Schloss, R. J., 2011. Building a highly consumable semantic model for smarter cities. *In Proceedings of the AI for an Intelligent Planet* (pp. 1-8).

Voitenko, G., 2023. Method for 3D neighborhood model creation in CityGML standard as basis for urban simulation tools. Master's thesis, Norwegian University of Science and Technology (NTNU), 70 p.